**Pig Configuration**

# 1. Download Apache

|  |
| --- |
| wget https://dlcdn.apache.org/pig/pig-0.17.0/pig-0.17.0.tar.gz |

A screenshot of a computer

Description automatically generated



# 2. Setup Pig

|  |
| --- |
| tar -xzf pig-0.17.0.tar.gz |



|  |
| --- |
| mv pig-0.17.0 pig |



# 3. Khai báo biến môi trường

|  |
| --- |
| nano ~/.bashrc |
| export PIG\_HOME=/home/hadoopngocphung/pig  export PIG\_CLASSPATH=$HADOOP\_CONF\_DIR  export PATH=$PATH:$PIG\_HOME/bin |

A screen shot of a computer program

Description automatically generated

# 4. Check Pig

|  |
| --- |
| pig -help |

A screenshot of a computer screen

Description automatically generated

# 5. Thực thi pig

Pig có 06 chế độ thực thi

- **Local Mode:** Chạy trên máy đơn. Tất cả các tập tin được cài đặt và chạy trên hệ thống file và máy chủ cục bộ.

|  |
| --- |
| pig -x local |

**- Tez Local Mode:** Tương tự chế độ cục bộ, ngoại trừ bên trong Pig sẽ gọi công cụ thực thi Tez.

|  |
| --- |
| pig -x tez\_local |

**- Spark Local Mode:** Tương tự chế độ cục bộ, ngoại trừ bên trong Pig sẽ gọi công cụ thực thi Spark.

|  |
| --- |
| pig -x spark local |

**- Map Reduce Mode:** Chạy ở chế độ mapreduce, cần truy cập cụm Hadoop và HDFS. Đây là chế độ chạy mặc định, có thể không cần cung cấp thông số -x

|  |
| --- |
| pig |

Hoặc

|  |
| --- |
| pig -x mapreduce |

**- Tez Mode:** Cần truy cập cụm Hadoop và hệ thống HDFS

|  |
| --- |
| pig -x tez |

**- Spark Mode:** Cần truy cập Spark, cụm Yarn hoặc Mesos và hệ thống file HDFS. Pig Script chạy trên Spark có thể tận dụng tính năng cấp phát động (dynamic allocation). Tính năng này có thể được kích hoạt bằng cách bật spark.dynamicAllocation.enabled.

|  |
| --- |
| pig -x spark |